CasperSent:
A Program for Computer-Assisted Speech Perception Testing and Training at the Sentence Level

Arthur Boothroyd
San Diego, California

The CasperSent program was designed mainly for auditory, visual, and auditory-visual speech-perception training at the sentence level. Design is informed by a model of speech perception with 4 main components. Two are derived from the speech stimulus, namely sensory evidence and contextual evidence. The other 2 are provided by the perceiver, namely knowledge and skill. The principal target of training with CasperSent is perceptual skill. This involves such things as attention control, use and balance of sensory and contextual evidence, balance of speed and accuracy, confidence, risk-tolerance, and error-handling. The software may also be used for performance assessment while measuring the effects of talker, perceptual modality, angle of view (if lipreading), topic knowledge, sentence length, sentence type, and, if used in combination with a clinical audiometer, signal-to-noise ratio. The software operates on a personal computer with a minimum of peripheral equipment. Testing and training can be self-administered, helper-assisted, or clinician-controlled. Stimuli consist of 720 sentences. Performance measures are based on sets of 12 sentences, and are automatically logged for later recall, graphing, and analysis. It has been shown that performance improves over time, that rate of improvement depends on the type of feedback, and that prior knowledge of sentence topic increases performance. In addition to its intended application in aural rehabilitation, CasperSent also has potential value in research and teaching.

The purpose of this paper is to describe a computer program designed as a tool for enhancing and measuring speech perception. Speech perception depends on

---

Arthur Boothroyd, Professor Emeritus, City University of New York.
Correspondence concerning this article should be addressed to Arthur Boothroyd, 2550 Brant Street, San Diego, California 92101.

learned relationships between patterns of neural activation in the auditory and visual systems and the language patterns from which these patterns originate (Boothroyd, 1997). These relationships change when there is a significant deterioration of hearing or a significant change in sensory assistance. Relearning, or adaptation, is then required (Neuman, 2005). To a certain extent relearning can occur as a result of self-hearing and participation in everyday communication. For many persons with hearing loss, however, the opportunities for every day communication are too few, the demands are too great relative to ability, or the consequences of error are too threatening. In such cases, relearning can be slow or non-existent and formal group or individual training is called for. Even then, constraints of cost and time can impose serious barriers on the amount and effectiveness of training. Developments in the capabilities and affordability of personal computers have led to the possibility of self training, or helper-assisted training at relatively low cost (Ross, 2005; Sweetow & Sabes, 2007). Examples include Listening and Communication Enhancement (LACE) from Neurotone,[1] which operates solely in the hearing domain and Seeing and Hearing Speech from Sensimetrics,[2] which includes lipreading.

The software described here is called CasperSent. Its main purpose is to improve speech perception at the sentence level. It does so, essentially, by increasing the total time the user spends on task. The program automatically tracks performance changes over time. It can also provide information on such things as the contributions of lipreading and topic knowledge, the effects of sentence length and type, and generalization to a novel talker. This information can have value in counseling. The original version of this software was developed in the 1980s for research on the rehabilitation of adult users of an early multi-channel cochlear implant (Boothroyd, 1987; Boothroyd, Hanin, & Waltzman, 1987). Prerecorded speech materials were played under computer control from 12 in. video laserdiscs and the software required clinician administration. The program described here has been adapted for use with a personal computer, using little or no peripheral equipment. The options of self-training and helper-assisted training have also been added, though the software can still be used under clinician control.[3]

A MODEL OF SPEECH PERCEPTION

The model of speech perception on which CasperSent is based assumes that the perceptual decisions required to translate a speech stimulus into language patterns rely on four main factors:

---

[1] www.neurotone.com

[2] www.sens.com

[3] CasperSent is one of a series of programs designed for training, testing, and studying speech perception at various levels of linguistic complexity. Other programs in this series deal with perception at the levels of phonetic contrasts, phonemes, and isolated words.

1. Sensory Evidence

Sensory evidence consists of patterns of neural stimulation generated by the perceiver's auditory and/or visual systems in response to the received speech stimulus. This stimulus originates from movements of the talker's speech mechanism. On the basis of the sensory evidence, the perceiver makes inferences about those movements and the language patterns they were intended to represent.

2. Contextual Evidence

Linguistic messages never occur in isolation but in context. This context is of three main types: world, social, and linguistic. It provides the perceiver with additional evidence with which to evaluate the sensory evidence and can have a major influence on success (Boothroyd, 2002; Boothroyd & Nittrouer, 1988).

3. Knowledge

The perceiver brings a lot of prior knowledge to the perceptual task. This knowledge can be implicit or explicit and, like the context, it is of three main types: world, social, and linguistic. Prior knowledge plays at least three roles in speech perception. First, it determines how much of the information in the context is actually available to the perceiver as contextual evidence. Second, it magnifies the value of the sensory evidence. Third, it is the source of possible decisions about the language, its meaning, and its significance.

4. Skill

The result of speech-perception skill is the ability to use sensory and contextual evidence to make inferences about their language source, at a rate determined by the talker, with an acceptably low probability of error, while leaving enough processing capacity to assess the implications of the message and, perhaps, formulate a response.

The first two factors are derived from the stimulus. The second two are what the perceiving brings to the process. This model is illustrated in Figure 1.

For the individual with sensorineural hearing loss, sensory evidence is limited by damage to the auditory system and by the limited visibility of speech movements. A principal goal of the *sensory* aspects of rehabilitation is to minimize the limitations by optimal selection, adjustment, and use of hearing aids, cochlear implants, and assistive technology, together with management of the acoustic environment. For face-to-face communication, control of the visual environment is also important.

As indicated earlier, contextual evidence is conditioned by the listener's world, social, and linguistic knowledge. This knowledge is probably not malleable over the short term, but contextual evidence can be controlled during training and testing, as can the language content and the resulting sound patterns.

The perceiver's perceptual skills and strategies are the principal targets of the

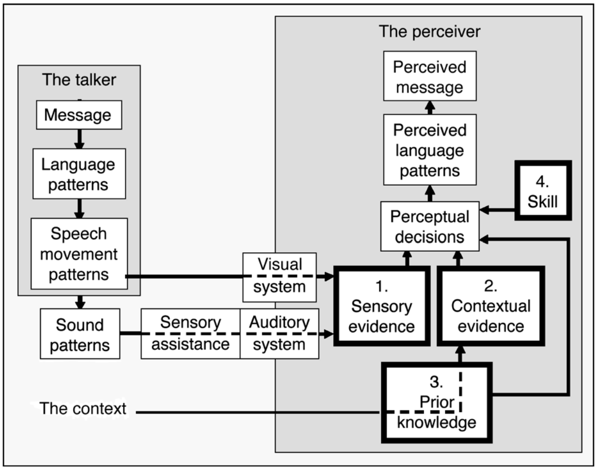

Figure 1. The model of speech perception that informs the design of the training and testing software described in this paper. Note that perceptual decisions are based on the four factors with bold outlines. The target of perceptual training is skill.

perceptual aspects of rehabilitation. In particular, we seek to enhance such things as attention, the use and balance of sensory and contextual evidence, perceptual speed, confidence, risk- and failure-tolerance, repair strategies, ease of listening, the balance between top-down and bottom-up processing, auditory-visual integration, and, ultimately, overall performance. There is also the possibility that perceptual rehabilitation can enhance the ability of the auditory system to generate useful sensory evidence – a process referred to as acclimatization (Arlinger et al., 1996; Cox, 1992; Gatehouse, 1993; Saunders & Cienkowski, 1997). And the role of vision in face-to-face communication cannot be ignored. Though limited, it provides a rich source of sensory evidence about speech movements and, therefore, about the underlying language patterns. For some listeners, this evidence supplements imperfect hearing. For others, vision is primary and hearing provides the supplement. And some must rely entirely on vision. Effective use of lipreading can be an important aspect of speech perception training and testing (Massaro & Cohen, 1999; Summerfield, 1983).

THE PROGRAM

The following sections provide information about the features and application of CasperSent.

1. Features

Equipment requirements. The program is designed to operate on a basic personal computer equipped with a sound card and amplified loudspeakers. When the software is under clinician control, the option exists for routing the audio signal through an audiometer. This option allows for calibrated level adjustment and the addition of noise. The computer must also have a CD drive for software installation and a DVD drive either for playing the speech files or for transferring them to the computer's hard drive. The software operates within Microsoft Windows. If installed on a Macintosh computer, the latter must be provided with appropriate software.[4]

Sentence materials. The program uses the CUNY topic-related sentence sets, originally developed at the City University of New York for studies of the rehabilitation of adult cochlear implant and tactile aid users. The sentences are intended to be representative of everyday conversation, and are organized into sets of 12. Each set contains 102 words.[5] An example of a sentence set is provided in the Appendix. There are 60 sentence sets, making it possible to provide fairly extended testing and training without repetition. The sentences in each set are related to 12 topics that remain constant from set to set. By giving the subject prior knowledge of sentence topic, one has the opportunity to more closely simulate a conversational context, in which the topic remains constant over a number of exchanges. Within each set, the number of words in a sentence varies uniformly from 3 through 14 words, making it possible to examine the influence of sentence length. In addition, each set contains four statements, four questions, and four commands, making it possible to examine the effect of sentence type. Length and type vary randomly across topic from set to set.

Recordings. At the time of writing, video recordings of three talkers are available for use with CasperSent. Two are women and one is a man. Talker #1 is filmed only at 0° azimuth but two sound tracks are available: (a) the full audio signal and (b) fundamental frequency (Fo) only. The latter is useful for the study and/or demonstration of the synergistic effects of lipreading and rudimentary hearing (Rosen, Fourcin, & Moore, 1981). The sound tracks for Talkers #2 and #3 are limited to the full audio signal but the video image can be viewed at three angles. These are 0, 45, and 90° azimuth (see Figure 2). This feature was added at the suggestion of a hearing-impaired consultant who pointed out that one is not guaranteed a 0° azimuth in real-life communication situations.

Self-administration. The testing and training can be self-administered, helper-assisted, or clinician-controlled. The second two applications require the addition of a second monitor.

Feedback. There are three options for feedback after a sentence has been re-

[4] See, for example, www.apple.com/macosx/features/bootcamp.html.

[5] Because of the sentence context, the 102 words cannot be treated as 102 statistically independent test items. Estimates of confidence limits for word recognition scores can, however, be derived from binomial probability theory if it is assumed that each set contains roughly 25 independent test items. The estimated 95% confidence limits for a 50% score based on only one set will then be +/-20 percentage points. This can be reduced to +/-10 percentage points by basing scores on the average of four sets.

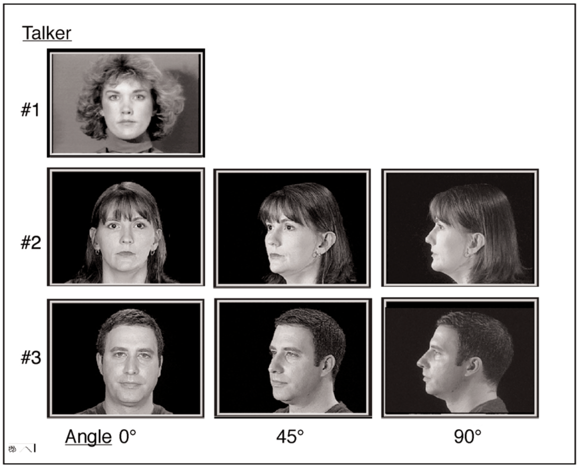

Figure 2. The three talkers incorporated into CasperSent.

peated: (a) none (for testing); (b) the full sentence text (the only option for self-administration); or (c) the text of the words correct, and the opportunity for a second try while focusing attention on words that were missed the first time. The last option is intended to accelerate learning while making more efficient use of the sentence materials.

Data logging. After each sentence set has been presented and scored, the results are automatically logged, on a drive of the user's choice, for later recall, analysis, and graphing. The logged data can be recalled within the program or within a spreadsheet or graphics program.

2. Hardware Setup

The simplest hardware setup is that used for self-administration via an amplified loudspeaker, as illustrated in the top panel of Figure 3. The user hears (and/or sees) a sentence and attempts to repeat it. He is then shown the text of the sentence and indicates which words were repeated correctly. He has an opportunity to hear (and/or see) the sentence again before moving on to the next one. The act of repetition is important for two reasons. First, it helps establish the relationship between the motor activities of speech and the resulting sensory evidence. Second, the user is more likely to score his initial perception accurately.

With the addition of a second monitor,[6] the scoring can be done by a second

[6]Lap-top computers automatically provide for the addition of a second monitor. Some desktop computers also have video cards that permit the use of two monitors. If not, the necessary dual output can be created with a simple splitter such as the VM-112A from the PTC group (http://thedigiview.com/SVGA.html).

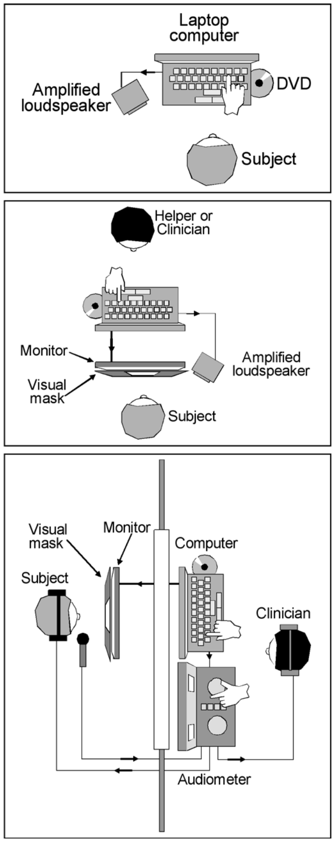

Figure 3. Hardware setup for administration of CasperSent. The upper panel shows a setup for self-administration. The middle panel shows a setup for helper-assistance. The bottom panel shows a setup for administration in an audiology clinic.

person, either a clinician or a helper, as illustrated in the middle panel of Figure 3. This option increases the objectivity of scoring and permits activation of the feedback routine in which the learner is shown only the words perceived correctly and has a second chance to process the sentence. Note the use of a simple cardboard mask to restrict the subject's screen view to the video image of the talker. This approach avoids the need for special hardware and/or software for independent operation of dual monitors.

When using the setups just described, sound presentation level can be calibrated using an inexpensive sound-level meter. If required, however, CasperSent can be administered via an audiometer, using either earphone or sound-field presentation, as illustrated in the bottom panel of Figure 3. In this case, the audio output of the computer provides input to the speech channels of the audiometer and the issues of calibration and presentation level are already addressed. An additional advantage is that testing and training can be carried out in noise.

The user can either leave the stimulus files on DVD or copy them to the computer's hard drive. If the first option is chosen, the DVD for the chosen talker and viewing angle must be in the DVD drive when the program is run. Copying the files to the hard drive avoids the necessity of having the appropriate DVD available but requires considerable storage capacity (roughly 6 GB for *each* of the three talkers).

3. Training and Testing

After entering the subject's and the tester's ID, selecting the drive containing the stimuli, and selecting the drive for the location of logged data, the user is directed to the training/testing screen. Here, there are several choices to be made:

1. The sentence set to be used (1 . . . 60)
2. The talker (T1/T2/T3; only those talkers whose recordings are found by the software are offered as choices)
3. The modality (Audio/Visual/Both)
4. For Talker 1, the audio signal (Full/Fo only)
5. For Talkers 2 and 3, the angle of view (0/45/90°)
6. Whether sentence topic will be shown or withheld
7. The type of feedback (None/Full text/words correct plus second try)

When "Begin" is clicked the first sentence in the selected set is presented under the specified condition. The topic (if selected) is also shown, and the subject is asked to repeat as much of the sentence as possible – guessing freely if not sure. The text of the sentence is then shown to the tester (or to the subject if self-administered) with numbers under each word. The response is scored by entering the numbers of the correct words on the keyboard or by clicking the correct words on the screen. The number of words correct is automatically recorded in the table at the top of the screen. Depending on the feedback mode, the subject is then shown nothing, the full text of the sentence, or the text of the sentence with missed words indicated by dashes. In the last case, he is given a second try and the new response is scored – the result, again, being recorded at the top of the screen. After all 12 sentences in a set have been presented, the data are automatically logged, together with any notes added by the tester. There is also the option to print the screen as insurance against possible data loss. The upper panel of Figure 4 shows the training/testing screen, as it would appear to a helper or clinician, during testing with "Partial and 2nd try" selected for feedback. The lower

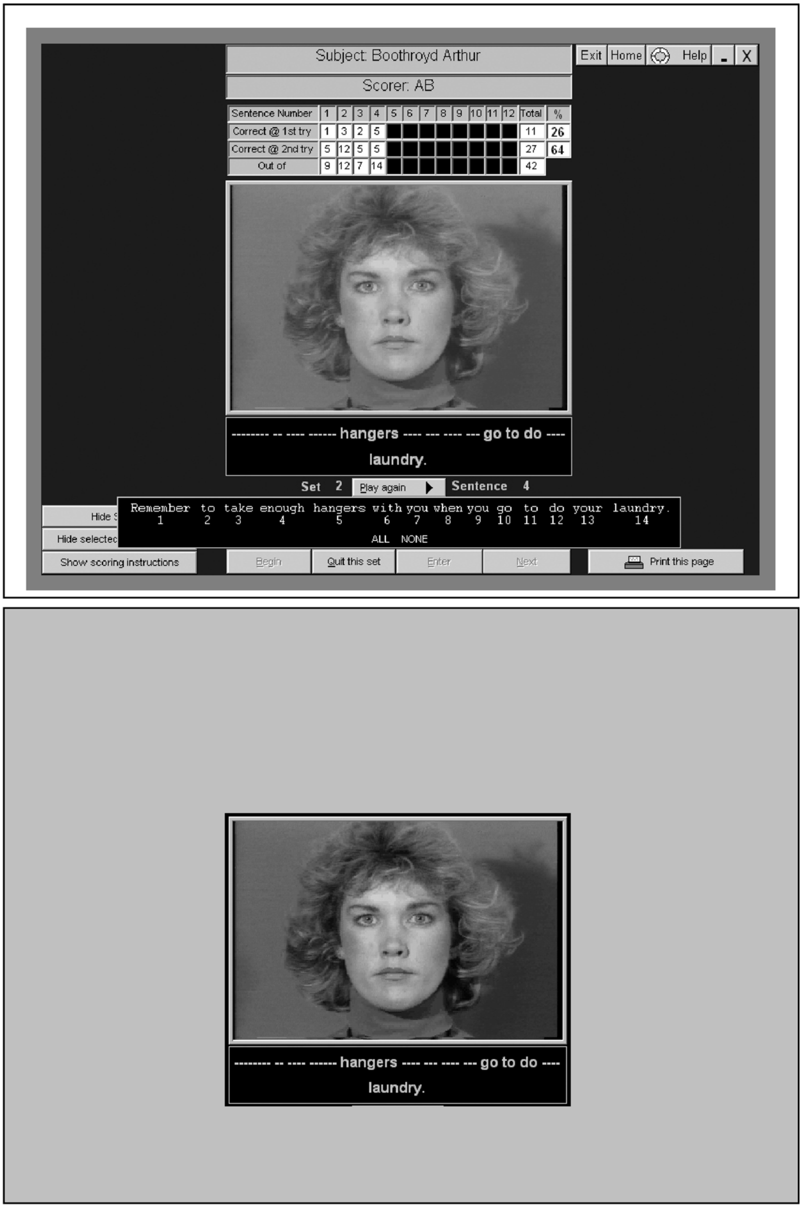

Figure 4. The testing screen. The upper panel shows the full screen. The bottom panel shows the portion of the screen seen by the subject during helper-assisted or clinician-controlled training. In this example, feedback to the subject shows words correct at the first try with dashes for the words still to be recognized.

panel shows the screen as it would appear to the subject. In self-administered training, the subject would, of course, see the complete screen.

4. Data Retrieval

Figure 5 shows a subject's record after it has been retrieved within the program. Shown on the left is information on the sentence sets presented, together with time, test conditions, and the percentage score(s). At the right is detailed information for the most recently selected record, showing the 12 test sentences, with correct words highlighted. At the bottom are the numbers of words correct in each sentence, shown in both tabular and graphic form. The button labeled "Show graphs" takes the user to a graphic display of the selected records, where scores are shown as a function of time, and also averaged across sessions.

SAMPLE DATA

1. Learning

The original version of CasperSent was used in several studies of the perceptual rehabilitation of users of tactile aids and cochlear implants, as well as in more

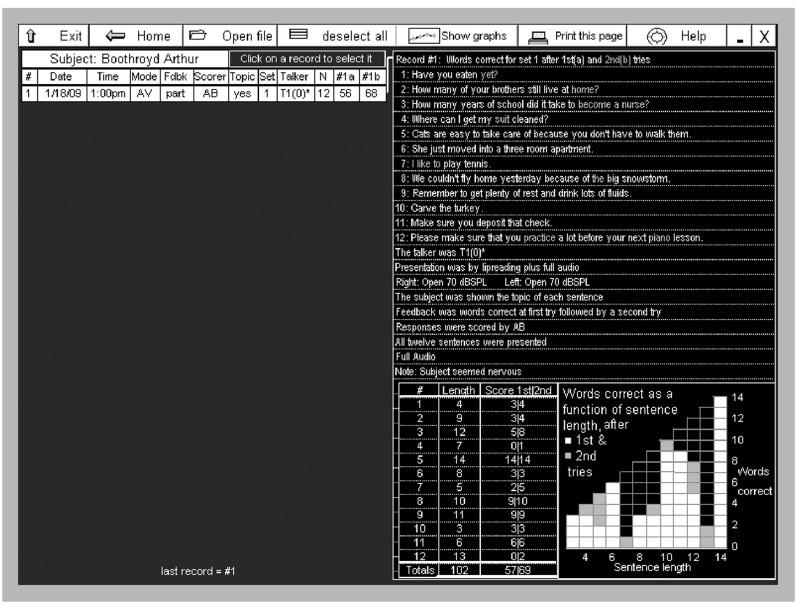

Figure 5. An example of subject data, as they are recalled and displayed within CasperSent.

basic research (Boothroyd et al., 1987; Boothroyd & Hnath-Chisolm, 1988; Hanin, 1988; Hnath-Chisolm & Boothroyd, 1992). Figure 6 shows an example of results from one of these studies in which eight adults with profound hearing loss were trained in the use of a spatial-tactile display of voice fundamental frequency as a supplement to lipreading (Boothroyd, Kishon-Rabin, & Waldstein, 1995). Blocks of training with lipreading plus the device were interspersed with training and testing by lipreading alone. The lines in Figure 6 are least-square fits to an exponential growth function:

$$y = a + (b - a)\left\{1 - \exp\left[\frac{-(x - 1)}{c}\right]\right\} \quad (1)$$

where:

x = number of sentence sets presented

y = percentage words recognized in set x

a = words recognized in the first set

b = asymptotic performance

c = a growth function inversely related to rate of learning

Mean scores increased during training by around 10 percentage points for lipreading alone but by around 20 percentage points for tactile enhancement. It could be argued that the subjects had more training with the device than by lipreading alone. But if the device was not providing benefit, then all of the training would have been by lipreading only. This was just one of many studies

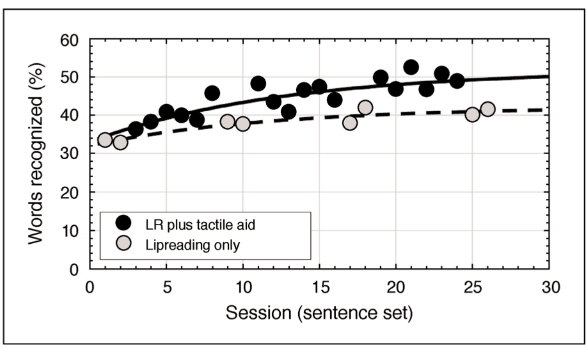

Figure 6. Learning data from a study of lipreading enhancement with a spatial tactile display of voice fundamental frequency. Data points are means (+/-1 standard error) for eight adults with profound hearing loss. From "Studies of Tactile Speechreading Enhancement in Deaf Adults," by A. Boothroyd, L. Kishon-Rabin, and R. Waldstein, 1995, *Seminars in Hearing, 16*, pp. 328-342. Used with permission.

of performance improvement over time, using the techniques embodied in CasperSent.

2. Effect of Feedback

During development of the original version of Casper at the City University of New York, a small pilot study (unpublished) was conducted to compare feedback options during sentence-level training. Fifteen normally hearing young adults were trained by lipreading alone. The subjects were divided into three groups. Group 1 repeated as much as they could of each sentence but received no feedback on performance. Group 2 was shown the text of each sentence after attempting to repeat it. Group 3 was shown the words correctly repeated and given a second and a third try. If there were still words missed at the third try, the full text was shown, with an additional opportunity to see the sentence spoken. We referred to this last procedure as "semi-automated tracking" (DeFilippo & Scott, 1978). It was more complex and time consuming than the "Partial text plus second try" incorporated into the revised version of CasperSent but had the same goals – to facilitate learning by increasing time on task while focusing attention on words giving the most difficulty. Group mean performance is shown as a function of sentence set in Figure 7. Error bars show standard errors estimated

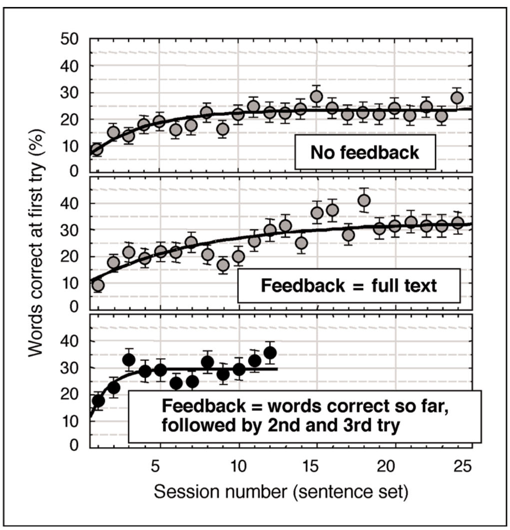

Figure 7. Words recognized by lipreading as a function of sentence set under three feedback conditions. Data points are means (+/-1 standard error) for groups of five inexperienced adult listeners with normal hearing. Lines are least-square fits to exponential growth functions. The third group appears to show a more rapid learning rate. When plotted in terms of time on task, however, these data show very similar learning rates.

from binomial theory by assuming that each score is based on 125 independent items (5 subjects by 25 items per set). The data have been fit by the simple exponential growth function of Equation 1.

At first sight, it appeared that the rate of learning was higher for the semi-automated tracking procedure. Bear in mind, however, that the subjects using this procedure spent roughly 3 times as long “on task” for each sentence set as did subjects in the other two groups. If the *x* axis is adjusted to represent estimated time on task, the rate of learning is similar for all three groups. These results suggest that the key determinant of learning rate is time spent on task rather than the amount of sentence material covered. In other words, the semi-automated tracking procedure makes more effective use of a limited amount of pre-recorded training material because of multiple repetitions. The results of this pilot study justified the retention of the semi-automated tracking procedure, albeit simplified, in the revised version of CasperSent.

Note that the amount of learning for groups 2 and 3 was somewhat greater than that for the no-feedback group. Unfortunately, the limited sample size and high within-group variability made it impossible to generalize this observation with any confidence.

3. Effect of Topic Knowledge

During ongoing studies of speech perception at San Diego State University, Mackersie has measured word recognition by 16 young normally hearing adults with and without information about sentence topic (Mackersie, Finnegan, & Boothroyd, 2006). Subjects were tested by lipreading alone and lipreading supplemented by the audio signal presented at 0, 3, 6, 9, and 12 dB relative to speech detection threshold. Group means (+/-1 standard error) are shown in Figure 8. The solid line is a least-square fit to the equation:

$$y = 100 * \left[1 - (1 - \left(\frac{x}{100}\right)^{k}\right] \qquad (2)$$

where:

y = percentage words recognized with topic knowledge

x = percentage words recognized without topic knowledge

k = a dimensionless factor representing the contribution of topic knowledge

The k-factor can be thought of as the amount by which knowledge of topic multiplies the effective information content of the sentences (Boothroyd & Nittrouer, 1988). The k-factor of 1.44 found here, applies only to these sentence materials and to this recording of this talker. It does, however, illustrate the importance of making a hearing-impaired conversational partner aware of topic changes. Measurement of the k-factor, by testing with and without topic, may help indicate how much use an individual subject is making of this valuable source of information.

4. Synergy of Lipreading and Voice Fundamental

The synergistic effect of lipreading and rudimentary hearing has been reported by many researchers (Breeuwer & Plomp, 1986; Grant, Ardell, Kuhl, & Sparks, 1985; Kishon-Rabin, Boothroyd, & Hanin, 1993; Risberg, 1974; Rosen et al., 1981). Some of this effect can be attributed to information on the presence or absence of voicing, but most depends on access to variations of fundamental frequency over time (Boothroyd, Waldstein, & Yeung, 1992; Hnath-Chisolm & Boothroyd, 1992; Waldstein & Boothroyd, 1994). The Fo-enhancement effect is well illustrated by some recent data obtained with CasperSent and shown in Figure 9. These data are taken from a class demonstration in which CasperSent was used to present sentences to 31 young adult students via Fo only, lipreading only, and the two in combination. One sentence set was presented under each condition and students wrote down as much as they could of the 36 sentences. Because this was a class demonstration, the sound was presented via loudspeaker and the visual image was presented on a large screen via an LCD projector. Moreover,

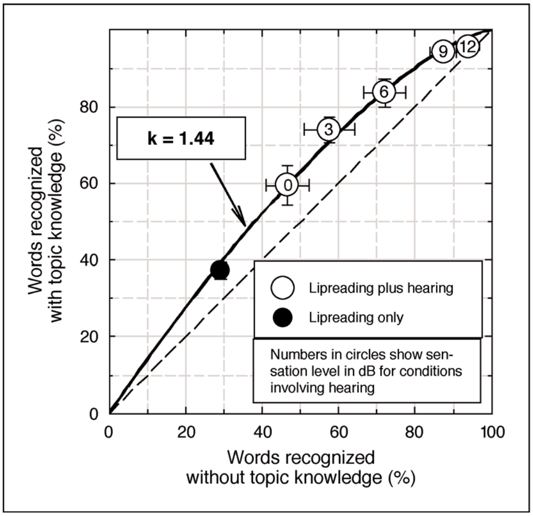

Figure 8. The effect of knowledge of sentence topic on the word recognition scores of 16 young normally hearing adults tested with CasperSent by lipreading alone and lipreading supplemented by hearing at various sensation levels. The heavy line shows a least-squares fit to Equation 2. The broken line shows the expected relationship if topic knowledge were providing no information. From "Sentence Perception With Impoverished Auditory Input: Effects of Lipreading and Topic-Knowledge," by C. Mackersie, D. Finnegan, & A. Boothroyd, 2006, unpublished manuscript. Used with permission.

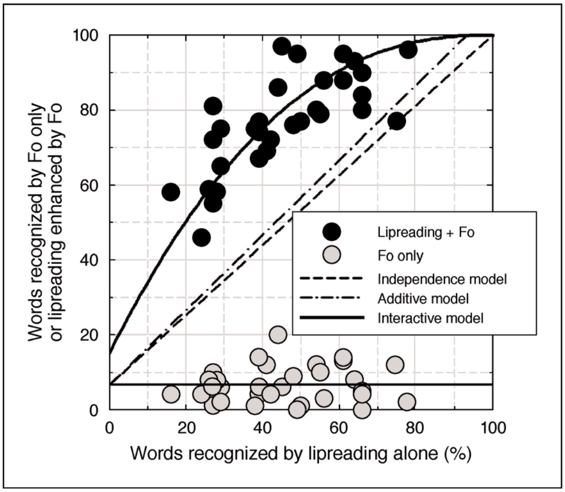

Figure 9. Words recognized in sentences when listening to fundamental frequency (Fo) only, and when lipreading is combined with Fo. Both are shown as a function of words recognized by lipreading only. Data points are for 31 young adults responding to a single sentence set under each condition. Broken lines show predictions for the combined score under an independence model and an additive model (see text). The solid curve is the least-square fit to an interactive model in which the effect of Fo is to multiply the information extracted from the visual stimulus by a factor of 2.4.

all students observed the same three sentence sets and experienced the three conditions in the same order. There is, therefore, the potential for confounding of modality with inter-set differences and order effects. Nevertheless, the data are in keeping with numerous reports from appropriately designed research studies.

On average, this group of students recognized 46% of the words by lipreading only. When Fo was presented alone, they recognized only 7% of the words. But when Fo was added to lipreading, the mean score rose from 46% to 77% – an increase of 31 percentage points. There was no evidence of an association between scores for lipreading alone and Fo alone. In other words, the better lipreaders did not necessarily score better with the rudimentary audio signal. The better lipreaders did, however, tend to be the better scorers when lipreading was enhanced by Fo.

What is remarkable about these data, and other data reported in the literature, is the marked synergy between the two sources of information. The Fo signal, virtually unintelligible by itself, dramatically increases performance when combined with lipreading. Moreover, the magnitude of this effect is considerably higher than can be predicted by assuming either that the visual and auditory channels simply provide independent access to word identity (in which case the error probability under the combined condition would be the product of the error probabilities under the separate conditions), or that the channels are complementary (in which case, the percentage score under the combined condition would be the sum of the scores under the two separate conditions). Predictions for these two models are shown by the broken lines in Figure 9. The enhancement effect can be modeled, however, if we assume *interaction* between the two sources of information – in other words, that the Fo information increases the utility of the visual image. This effect can be modeled by Equation 2, as shown by the full heavy curve in Figure 9. The value of k providing a least-squares fit to the data was 2.4. In other words, the auditory stimulus, which is unintelligible by itself, multiplies the effective information in the visual stimulus by a factor of 2.4 (a 140% increase). To put these data into perspective, highly competent lipreaders typically score in excess of 70% on the CUNY sentences when presented by lipreading alone. In the data reported in Figure 9, only 2 of the 31 students scored better than 70% by lipreading only, but over two thirds exceed this score when the Fo signal was available.

DISCUSSION

As indicated in the introduction, the goal of perceptual rehabilitation is improvement of speech perception skill and/or modification of perceptual strategy. Programs such as CasperSent can pursue this goal by: (a) increasing time on task; (b) providing consistent feedback on performance; (c) developing associations among textual, auditory, visual, and motoric representations of speech; and (d) minimizing of the negative consequence of error. To the extent that these programs can be used by the hearing-impaired person alone, or with the help of a friend or significant other, the associated costs are small. In theory, they offer the potential for accelerated learning and improved outcome, in a cost-effective manner, when incorporated into a program of sensory and perceptual rehabilitation. They should not, however, be seen as a substitute for every day communication, but as a supplement.

One asset of computer-based training is that it provides automatic documentation of performance and progress. These data, however, are only useful to the extent that they reflect performance and progress in every day communication. It is relatively easy to demonstrate learning effects using software such as CasperSent. The challenge is to determine the true source and nature of this learning. Much of the change over time shown in Figures 6 and 7, for example, could reflect growing familiarity and comfort with the equipment, the task, the talker, and the type of sentence material. How much of the change represents generalizable improvements in underlying speech-perception skills or strategies is uncertain. There is, in fact, a pressing need for research into the efficacy of training that focuses on the enhancement of speech perception skills, regardless of whether such training involves computer-assistance. At the very least, such research should assess carry-over to different talkers, different materials, and realworld communication. Ideally it would also include measures of ease-of listening, stress, perceptual speed, and confidence. Such work should be supplemented by qualitative and grounded-theory research using both self-report and focus groups.

Note, also, that a measure of word recognition in sentences, while it may provide a useful index of function, provides little analytic or diagnostic information. It does not, for example, indicate which aspects of the perceptual process outlined in Figure 1 are most responsible for communicative failure. As a result, it does not identify components of the process that are most in need of attention. To meet this need one must also measure performance at the phonetic, phonemic, and isolated word levels (Hanin, 1988).

The issue of growing familiarity with the type of test material was mentioned above. A characteristic of the CUNY topic-related sentences is the frequent use of phrases such as "Please remember to . . .," "Remember to . . .," and "Don't forget to . . . ." When the materials were first developed, there was a tendency to pad sentences with such phrases in order to meet the stringent criteria for distribution of sentence length within a set. Unfortunately, this feature created an unforeseen source of predictability that can contribute to non-generalizable short-term learning effects. When generating new materials care should be taken to avoid this effect.

At the time of writing there is considerable interest in training by hearing alone. Much of this interest comes from clinicians working with cochlear implant recipients. Although the current version of CasperSent can be used for this purpose, the demands for computer memory would be dramatically reduced by having a parallel audio-only version. At the same time, it would be much easier to allow users to add new sentence materials and new talkers than is the case with the current audio-visual version. Such a feature would help address the need for studies of carry over.

Hearing aid and implant users with good sentence-level skills in quiet, often have considerable difficulty in the presence of noise – especially when this comes from one or more competing talkers. The present version of CasperSent does not, by itself, allow for training or testing in the presence of noise. This feature is easily added, however, if an audiometer, or other presentation system, is used, as in the bottom panel of Figure 3. Unfortunately, the type and level of noise will not be logged automatically, but must be entered manually by the tester. Future versions of CasperSent will incorporate the option for training and testing in noise.

SUMMARY

CasperSent is a multi-media program, designed for the study, measurement, and improvement of sentence-level speech perception by persons with hearing loss. By increasing time on task in a non-threatening context, with feedback on performance and multiple opportunities for repetition, this and similar programs have the potential to enhance speech perception skills in ways that will carry over to every day communication. There is a pressing need, however, to confirm that this potential is realized. The avoidance of custom-built peripheral equipment and the incorporation of self- or helper-assisted administration add the benefit of low cost. Automatic data logging provides documentation of performance and progress and may contribute to evidence of efficacy for this approach to perceptual rehabilitation. CasperSent can also be used for research, and for demonstration to students, clinicians, and significant others.

ACKNOWLEDGEMENTS

The contributions of numerous former and current colleagues are gratefully acknowledged. These include: Theresa Hnath-Chisolm, Laurie Hanin, Eddy Yeung, Liat Kishon-Rabin, Orna Eran, Larry Medwetsky, Robin Waldstein, Cecil Redmond, Gary Chant, Nancy Plant, Carol Mackersie, and Brianne Burger. The original development of CASPER was funded by NIH grant # 2P50DC00178 to the City University of New York. Current work on the CasperSent revision is funded by NIDRR grant #H1343E98 to the Rehabilitation Engineering Rehabilitation and Research Center at Gallaudet University.

AVAILABILITY

The CasperSent software is available for evaluation. It can be downloaded, together with the instruction manual and a small sample of the stimuli, from www.arthurboothroyd.com. Persons interested in acquiring the full stimulus set on DVD should contact Dr. Matthew Bakke, director of the Gallaudet University Rehabilitation Engineering Rehabilitation and Research Center at matthew.bakke@gallaudet.edu.

REFERENCES

Arlinger, S., Gatehouse, S., Bentler, R.A., Byrne, D., Cox, R.M., Dirks, D.D., et al. (1996). Report of the Eriksholm workshop on auditory deprivation and acclimatization. *Ear and Hearing, 17,* 87s-98s.

Boothroyd, A. (1987). *CASPER: Computer-Assisted speech perception evaluation and training*. Paper presented at the 10th annual Conference on Rehabilitation Technology, Washington, DC.

Boothroyd, A. (1997). Auditory development of the hearing child. *Scandinavian Audiology, 26*(Suppl. 46), 9-16.

Boothroyd, A. (2002). *Influencia del contexto en la percepción del lengaje hablado*. Paper presented at the Congreso Internacional de Foniatría, Audiología, Logopedia y Psicología del lenguaje, Salamanca.

Boothroyd, A., Hanin, L., & Waltzman, S.B. (1987). *Development of speech perception skills in cochlear implantees*. Paper presented at the 10th annual Conference on Rehabilitation Technology, Washington, DC.

Boothroyd, A., & Hnath-Chisolm, T. (1988). Spatial, tactile presentation of voice fundamental frequency as a supplement to lipreading: Results of extended training with a single subject. *Journal of Rehabilitation Research and Development, 25,* 51-56.

Boothroyd, A., Kishon-Rabin, L., & Waldstein, R. (1995). Studies of tactile speechreading enhancement in deaf adults. *Seminars in Hearing, 16,* 328-342.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America, 84,* 101-114.

Boothroyd, A., Waldstein, R., & Yeung, E. (1992). *Investigations into the auditory Fo speechreading enhancement effect using a sinusoidal replica of the Fo contour.* Paper presented at the International Conference on Spoken Language Processing, Banff, Canada.

Breeuwer, M., & Plomp, R. (1986). Speechreading supplemented with auditorily presented speech parameters. *Journal of the Acoustical Society of America, 79,* 481-489.

Cox, R. (1992). Maturation of hearing aid benefit: Objective and subjective measurements. *Ear and Hearing, 13,* 131-141.

DeFilippo, C.L., & Scott, B.L. (1978). A method for training and evaluating the reception of ongoing speech. *Journal of the Acoustical Society of America, 63,* 1186-1192.

Gatehouse, S. (1993). Role of perceptual acclimatization in the selection of frequency responses for hearing aids. *Journal of the American Academy of Audiology, 4,* 296-306.

Grant, K., Ardell, L., Kuhl, P., & Sparks, D. (1985). The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *Journal of the Acoustical Society of America, 77,* 671-677.

Hanin, L. (1988). *Effect of experience and linguistic context on speechreading*. New York: City University of New York.

Hnath-Chisolm, T., & Boothroyd, A. (1992). Speechreading enhancement by fundamental frequency: The effects of Fo contour distortions. *Journal of Speech and Hearing Research, 35,* 1160-1168.

Kishon-Rabin, L., Boothroyd, A., & Hanin, L. (1993). Speechreading enhancement using a spatial-tactile display of voice fundamental frequency (Fo): Comparison with auditory Fo and cochlear implants. *Journal of the Acoustical Society of America, 100,* 593-602.

Mackersie, C., Finnegan, D., & Boothroyd, A. (2006). *Sentence perception with impoverished auditory input: Effects of lipreading and topic knowledge*. Unpublished manuscript.

Massaro, D.W., & Cohen, M.M. (1999). Speech perception in perceivers with hearing loss: Synergy of multiple modalities. *Journal of Speech, Language and Hearing Research, 42,* 21-41.

Neuman, A.C. (2005). Central auditory system plasiticity and aural rehabilitation of adults. *Journal of Rehabilitation Research and Development, 42*(4), 169-186.

Risberg, A. (1974). The importance of prosodic speech elements for the lipreader. *Scandinavian Audiology, 4,* 153-164.

Rosen, S.M., Fourcin, A.J., & Moore, B.C.J. (1981). Voice pitch as an aid to lipreading. *Nature, 291,* 150-152.

Ross, M. (2005). Home-based auditory and speechreading training. *Hearing Loss, 26*(6), 30-34.

Saunders, G.H., & Cienkowski, K.M. (1997). Acclimatization to hearing aids. *Ear and Hearing, 18,* 129-139.

Summerfield, A.Q. (1983). Audio-visual speech perception. In M.E. Lutman & E. Hagberg (Eds.), *Hearing science and hearing disorders* (pp. 132-179). New York: Academic Press.

Sweetow, R.W., & Sabes, J.H. (2007). Technological advances in aural rehabilitation: Applications and innovative methods of delivery. *Trends in Amplification, 11,* 101-111.

Waldstein, R.S., & Boothroyd, A. (1994). Speechreading enhancement using a sinusoidal substitute for voice fundamental frequency. *Speech Communication, 14,* 303-312.

APPENDIX

ONE OF THE CUNY TOPIC-RELATED SENTENCE SETS, LISTED BY TOPIC, LENGTH, AND TYPE

Number	Topic	Length	Type[a]	
1	Food	9	Q	Do you want to have a barbecue this evening?
2	Family	12	Q	When was the last time that you went to visit your parents?
3	Work	7	Q	When will you be taking your vacation?
4	Clothes	14	C	Remember to take enough hangers with you when you go to do your laundry.
5	Animals & Pets	8	S	You can see deer near my country house.
6	Homes	6	S	We are looking for an apartment.
7	Sports & Hobbies	10	S	The football field is right next to the baseball field.
8	Weather	11	Q	Can you remember the last time we had so much snow?
9	Health	3	C	See your doctor.
10	Seasons & Holidays	5	C	Put these lights on the tree.
11	Money	13	C	Don't use your credit card if you can't pay the bill on time.
12	Music	4	S	I like that song.

[a] Q = Question; S = Statement; C = Command